Neural Network Applications for NWP Models

V. Krasnopolsky

Some NN applications in Numerical Modeling

• **Model Initialization:**
  – Satellite Retrievals
  – Direct Assimilation
  – Assimilation of Surface Observations

• **Model Physics:**
  – Fast and Accurate NN Emulations of Model Physics
  – New NN parameterizations
  – NN Stochastic physics

• **Post-processing:**
  – NN nonlinear ensembles
  – NN nonlinear bias corrections
  – Upscaling and downscaling
NN Applications Developed (black) and Under Development (red)

• **NN for Model Initialization:**
  – Satellite Retrievals
    • SSMI retrieval algorithm (operational since 1998)
    • QuickScat retrieval algorithm
  – Direct Assimilation
    • Forward model for direct assimilation of SSMI BT
    • QuickScat forward model
  – Assimilation of Surface Observations
    • Observation operator for assimilation of SSH anomaly
    • Empirical biological model for ocean color
    • NN algorithm to fill gaps in ocean color fields and for creating long and consistent ocean color data sets
NN Applications Developed (black) and Under Development (red) – cont.

• **NN for Model Physics:**
  – Fast and accurate emulations of parameterizations of physics
    • Fast nonlinear wave-wave interaction for WaveWatch
    • Fast NN long and short wave radiation for NCAR CAM, CFS, and GFS models
    • NN emulation for CRM in MMF
    • Fast NN microphysics for NMMB and WARF
  – New parameterization
    • Convection parameterization for NCAR CAM learned by NN from CRM
NN Applications Developed (black) and Under Development (red) – cont.

- **NN for Post-processing:**
  - Nonlinear ensembles
    - Nonlinear multi-model NN ensemble for calculating precipitation rates over ConUS
    - Nonlinear NN averaging of wave models ensemble
  - Nonlinear bias corrections
    - Nonlinear bias corrections for GFS
    - Nonlinear bias corrections for wave model
Mapping

• Mapping: A rule of correspondence established between vectors in vector spaces $^{n}$ and $^{m}$ that associates each vector $X$ of a vector space $^{n}$ with a vector $Y$ in another vector space $^{m}$.

\[ Y = F(X) \]
\[ X = \{x_1, x_2, \ldots, x_n\}, \quad ^n \]
\[ Y = \{y_1, y_2, \ldots, y_m\}, \quad ^m \]
\[ y_1 = f_1(x_1, x_2, \ldots, x_n) \]
\[ y_2 = f_2(x_1, x_2, \ldots, x_n) \]
\[ \vdots \]
\[ y_m = f_m(x_1, x_2, \ldots, x_n) \]
Mapping $Y = F(X)$: examples

- **Time series prediction:**
  \[ X = \{x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-n}\}, \text{ - Lag vector} \]
  \[ Y = \{x_{t+1}, x_{t+2}, \ldots, x_{t+m}\} \text{ - Prediction vector} \]
  (Weigend & Gershenfeld, “Time series prediction”, 1994)

- **Nonlinear ensemble average:**
  \[ X = \{\text{Ensemble members, Metadata}\} \]
  \[ Y = \{\text{Nonlinear ensemble average}\} \]
  (Krasnopolsky and Lin, 2012)

- **Retrieving surface wind speed over the ocean from satellite data (SSM/I):**
  \[ X = \{\text{SSM/I brightness temperatures}\} \]
  \[ Y = \{W, V, L, SST\} \]
  (Krasnopolsky, et al., 1999; operational since 1998)

- **Calculation of long wave atmospheric radiation:**
  \[ X = \{\text{Temperature, moisture, O}_3, \text{CO}_2, \text{cloud parameters profiles, surface fluxes, Metadata}\} \]
  \[ Y = \{\text{Heating rates profile, radiation fluxes}\} \]
  (Krasnopolsky et al., 2005, 2010, 2012)
**NN - Continuous Input to Output Mapping**

**Multilayer Perceptron**: Feed Forward, Fully Connected

\[ Y = F_{NN}(X) \]

\[ Y = \sum_{j=1}^{k} a_{qj} t_j = \sum_{j=1}^{k} a_{qj} (b_{j0} + \sum_{i=1}^{n} b_{ji} x_i) \]

\[ = a_{q0} + \sum_{j=1}^{k} a_{qj} \tanh(b_{j0} + \sum_{i=1}^{n} b_{ji} x_i) \]

\[ q = 1, 2, \ldots, m \]
I. NN in Model Initialization

Data Assimilation System (DAS)

"Ground" Observations

Satellite Data

Model Predictions

Initial Conditions

Oceanic (Atmospheric) Climate/Weather Prediction Numerical Model
Ingesting Satellite Data in DAS

- Satellite Retrievals:
  \[ G = f(S), \]
  \( S \) – vector of satellite measurements;
  \( G \) – vector of geophysical parameters;
  \( f \) – transfer function or retrieval algorithm

- Direct Assimilation of Satellite Data:
  \[ S = F(G), \]
  \( F \) – forward model

- Both \( F \) & \( f \) are mappings and NN are used
  - Fast and accurate NN retrieval algorithms \( f_{NN} \)
  - Fast NN forward models \( F_{NN} \) for direct assimilation
Wind speed fields retrieved from the SSM/I measurements for a mid-latitude storm. Two passes (one ascending and one descending) are shown in each panel. Each panel shows the wind speeds retrieved by (left to right) GSW (linear regression) and NN algorithms. The GSW algorithm does not produce reliable retrievals in the areas with high level of moisture (white areas). NN algorithm produces reliable and accurate high winds under the high level of moisture. 1 knot ≈ 0.514 m/s
DAS: Propagation of Information Vertically Using NNs

"Ground" Observations (mainly 2D) → \( \text{NN}_1 \) → Satellite Data (2D) → \( \text{NN}_2 \) → Ocean DAS

Model Predictions (3D & 2D)

NN1 and NN2 – observation operators
Observation operator for SSH

\[ Y = F(X) \approx NN(X), \]

\( Y \) – vector of SSH satellite measurements;

\( X \) – vector of ocean model variables;

\( F \) – observation operator – mapping

<table>
<thead>
<tr>
<th>Input #</th>
<th>Names</th>
<th>Units</th>
<th>Input</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Day of the year</td>
<td></td>
<td>( \sin\left(\frac{2 \cdot \text{day} - \pi}{366}\right) )</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Lon</td>
<td>( \sin(\text{lon}) )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Lon</td>
<td>( \cos(\text{lon}) )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Lon</td>
<td>( \sin(\text{lat}) )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>bottom pressure</td>
<td>?</td>
<td>pbot</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>vertically average x-velocity</td>
<td>cm/s</td>
<td>ubavg</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>vertically average y-velocity</td>
<td>cm/s</td>
<td>vbavg</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>temperature</td>
<td>°C</td>
<td>temp</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>layer thickness at p-points</td>
<td>m</td>
<td>dp</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>potential density</td>
<td>kg/m^3</td>
<td>th3d</td>
<td>32</td>
</tr>
<tr>
<td>11</td>
<td>internal x-velocity</td>
<td>cm/s</td>
<td>u</td>
<td>32</td>
</tr>
<tr>
<td>12</td>
<td>internal y-velocity</td>
<td>cm/s</td>
<td>v</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output #</th>
<th>Names</th>
<th>Units</th>
<th>Output</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SSH anomaly</td>
<td>m</td>
<td>srfhgt</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Montgomery potential</td>
<td>m^2/s^2</td>
<td>montg1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>
Profiles of derivatives of SSH. Derivatives over $dp$ (upper row), $temp$ (second row from the top), $th3d$ (third row from the top), $u$ (fourth row from the top) and $v$ (bottom row).
II. NN for Model Physics

- Deterministic First Principles Models, 3-D Partial Differential Equations on the Sphere + the set of conservation laws (mass, energy, momentum, water vapor, ozone, etc.)

\[ \frac{\partial \psi}{\partial t} + D(\psi, x) = P(\psi, x) \]

- \(\psi\) - a 3-D prognostic/dependent variable, e.g., temperature
- \(x\) - a 3-D independent variable: \(x, y, z\) & \(t\)
- \(D\) - dynamics (spectral or gridpoint)
- \(P\) - physics or parameterization of physical processes (1-D vertical r.h.s. forcing) – mostly time consuming part

> 50% of total time
Accurate and Fast NN Emulations of Model Physics

- Any **parameterization** of model physics is a mapping and can be emulated by NN.
- The **entire model physics** is a mapping and can be emulated by NN.
- NN emulation is usually **1 to 2 orders of magnitude faster** than the original parameterization: ~20 times for SWR and ~100 times for LWR.
The Magic of NN Performance

Input/Output Dependency: \[ Y = F(X) \]

Numerical Scheme for Solving Equations

Mathematical Representation of Physical Processes

\[ F(p) = B(p) \left( (p, p) + \int p \ (p, p) \ dB(p) \right) \]
\[ F(p) = B(p) \int p \ (p, p) \ dB(p) \]
\[ B(p) = T^4(p) \text{ the Stefan Boltzman relation} \]

\[ y_q = a_{q0} + k_{qj} \tanh(b_{j0} + \sum_{i=1}^{n} b_{ji} x_i), \quad q = 1, \ldots, m \]

Input/Output Dependency: \[ \{X_i, Y_i\}_{i=1}^{N} \]
NN Emulations of Model Physics Parameterizations

*Learning from Data*

- **Model**: $X \rightarrow F_{NN} \rightarrow Y$
- **Parameterization**: $X \rightarrow F \rightarrow Y$

*Training Set*: $\{X_i, Y_i\}, \ldots$ \quad $\forall X_i \in \mathbb{C}_{phys}$

**NN Emulation**: Final test of $F_{NN}$ in a parallel run of models with the physically based $F$ and with $F_{NN}$
LWR Individual Profiles

PRMSE = 0.18 & 0.10 K/day

PRMSE = 0.11 & 0.06 K/day

PRMSE = 0.05 & 0.04 K/day
2017

V. Krasnopolsky. NN Applications in NWP Models

CTL

NN FR

NN - CTL

CTL_O – CTL_N

JJA

NCEP CFS PRATE – 17 years

1 2 3 4 5 6 8 10 12 14 16 18 20

1 2 3 4 5 6 8 10 12 14 16 18 20

-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7

-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7
Anomaly correlation for temperature fields at 500 mb for the northern hemisphere (upper row), tropics (medium row), and southern hemisphere (lower row).

Two GFS (T574L64) runs are shown: black line – control run with the original LWR and SWR and red line – run with NN SWR and LWR.
III. Post-processing Model Output

- Nonlinear multi-model ensembles for precipitation forecast.
- Precipitation forecasts available from 8 operational models:
  - NCEP's mesoscale & global models (NAM & GFS)
  - the Canadian Meteorological Center regional & global models (CMC & CMCGLB)
  - global models from the Deutscher Wetterdienst (DWD)
  - the European Centre for Medium-Range Weather Forecasts (ECMWF) global model
  - the Japan Meteorological Agency (JMA) global model
  - the UK Met Office (UKMO) global model

- Also NCEP Climate Prediction Center (CPC) precipitation analysis is available over ConUS.
Calculating Ensemble Mean

- Conservative ensemble:
  \[ EM = \frac{1}{N} \sum_{i=1}^{N} p_i, \quad p_i \text{ is an ensemble member} \]

- Weighted ensemble:
  \[ WEM = \frac{\sum_{i=1}^{N} W_i p_i}{\sum_{i=1}^{N} W_i} \]

  \( W_i \) - from a priori considerations or from past data (linear regression)

- If past data are available the assumption of linear dependency can be relaxed:
  \[ NEM = f(P) \approx NN(P) \]
Sample NN forecast: example 1

Verification CPC analysis

NN

MEDLEY

HPC
Sample NN forecast: example 2

Verification CPC analysis

CPC 1/8 deg Analysis 24h Accum (mm) Ending 2010120612

MEDLEY 024h Forecast 24h Accum (mm) Ending 2010120612

NN 024h Forecast 24h Accum (mm) Ending 2010120612

HPC 024h Forecast 24h Accum (mm) Ending 2010120612
Sample NN forecast: example 3
Nonlinear Bias correction

• **Current approaches:**
  – Relate observed weather elements (**PREDICTANDS**) to appropriate model variables (**PREDICTORS**) via a statistical approach.

• **Predictors are obtained from:**
  1. Numerical Weather Prediction (NWP) Model Forecasts
  2. Prior Surface Weather Observations
  3. Geoclimatic Information

• **Predictands are obtained from:**
  – Historical record of observations at forecast points
Methodology

• **Statistical Method:**
  
  – **MULTIPLE LINEAR REGRESSION**
    
    \[ y_i = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \]
    
    \( y_i \) - predictand; \( i = 1, 2, \ldots, m \)
    
    \( X = \{ x_1, x_2, \ldots, x_n \} \) – vector of predictors
    
    FOR EACH SITE and FOR EACH PREDICTAND and FOR EACH WEATHER REGIME!

• **Motivations:**
  
  – Actual dependence \( y_i = f(X) \) may be nonlinear!
  
  – NNs are a flexible nonlinear tool that has a potential to represent these nonlinear dependencies more efficiently and compact
Data used in this study:

- observations for **two predictands** variables $T_{2m}$ and $T_d$
- collected at **3,000 stations over the ConUS**
- during the period of **669 days** starting June 1, 2013 and ending March 31, 2015.
- GFS 24h forecast was saved for **five predictors**, GFS variables $T_{2m}$, $T_d$, $rh$, $T_h$, and $wd$ during the same period of time. These variables plus some metadata (see below) were used as predictors.
- All days and locations with missing data are removed from the data sets.
- First 365 days were used for NN training, and last 304 days for independent validation
- Only 24 h projection time was considered
### Some results

**NN for all (3,000) stations, two variables (NN – 10:3:2)**

<table>
<thead>
<tr>
<th>in#</th>
<th>in</th>
<th>out</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Sind(jd)</td>
<td>T2mo<a href="jd+1">0</a></td>
</tr>
<tr>
<td>1</td>
<td>Cosd(jd)</td>
<td>Tdo<a href="jd+1">0</a></td>
</tr>
<tr>
<td>2</td>
<td>Lat</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Lon</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>GFS - T2m(jd)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>GFS – Td(jd)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>GFS – rh(jd)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>GFS – Th(jd)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>GFS – wd(jd)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>elevation</td>
<td></td>
</tr>
</tbody>
</table>

**NN performance, comparison with Linear regression bias correction**

<table>
<thead>
<tr>
<th></th>
<th>(T_{2m}) in °F</th>
<th>(T_d) in °F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td><strong>OBS - GFS</strong></td>
<td>1.38</td>
<td>5.16</td>
</tr>
<tr>
<td><strong>Linear Regression</strong></td>
<td>-0.081</td>
<td>4.84</td>
</tr>
<tr>
<td><strong>Best single NN</strong></td>
<td>0.068</td>
<td>4.11</td>
</tr>
<tr>
<td><strong>Ensemble of NN</strong></td>
<td>-0.078</td>
<td>4.09</td>
</tr>
</tbody>
</table>
Some results (cont.)

![Graph showing T2m bias in F as a function of elevation in feet. There are four lines representing different datasets: GFS - OBS, LR - OBS, Single NN - OBS, and NN ensemble - OBS. The graph indicates variations in bias with elevation.]
Conclusions

• NN is a generic and versatile AI technique
• There exist numerous applications in Numerical Weather Prediction that can be successfully approached using NNs
• In NWP models NNs can be used in model initialization, as parts of the model physics, and for post-processing model outputs
• A significant experience in developing NN applications for NWP models has been accumulated at EMC
There is no free lunch

• NN, as any statistical model, requires data for training
• NN, as any nonlinear statistical model, requires more data, than linear model/regression
• As any nonlinear statistical model, NN may be over fitted
• As any statistical model, NN should be periodically updated to changes in environment; NN can be updated on-line